Image Classification Using a Variational Approach

Christophe Samson, Laure Blanc-Féraud, Gilles Aubert, Josiane Zerubia

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Image Classification Using a Variational Approach

Christophe Samson — Laure Blanc-Féraud — Gilles Aubert — Josiane Zerubia

N° 3523

Octobre 1998

THÈME 3

ISSN 0249-6399

Image Classification Using a Variational Approach

Christophe Samson, Laure Blanc-Féraud, Gilles Aubert*, Josiane Zerubia

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 3523 — Octobre 1998 — 48 pages

Abstract: Herein, we present a variational model devoted to image classification coupled with an edge-preserving regularization process. The discrete nature of classification (i.e. to attribute a label to each pixel) has led to the development of many probabilistic image classification models, but rarely to variational ones. In the last decade, the variational approach has proven its efficiency in the field of edge-preserving restoration. In this paper we add a classification capability which contributes to provide images compound of homogeneous regions with regularized boundaries, a region being defined as a set of pixels belonging to the same class. The soundness of our model is based on the works developed on the phase transitions theory in mechanics. The proposed algorithm is fast, easy to implement, and efficient. We compare our results on both synthetic and satellite images with the ones obtained by a stochastic model using a Potts regularization.

Key-words: Variational model, classification, labelling, phase transitions theory, edge-preserving regularization, minimization, satellite images.

* Laboratoire J.A. dieudonné UMR 6621 CNRS, Université de Nice-Sophia Antipolis, 06108 Nice Cedex 2, FRANCE

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 —Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

Classification d'Images par Approche Variationnelle

Résumé : Dans ce rapport nous présentons un modèle variationnel destiné à la classification d'images avec processus de régularisation préservant les contours. La notion de classification étant par nature discrète (i.e. attribuer un label à chaque pixel de l'image), il existe de nombreux modèles de classification par approche probabiliste, mais les modèles variationnels abordant ce sujet sont rares. Ces dernières années, l'approche variationnelle a montré sont efficacité dans le cadre de la restauration d'images avec prise en compte des discontinuités. Dans ce travail, nous ajoutons un processus de classification permettant d'obtenir une solution formée de régions homogènes dont les frontières sont régulières (une région étant définie par l'ensemble des pixels appartenant à la même classe). La justification théorique de notre modèle repose sur les travaux effectués dans le cadre des problèmes de transitions de phases en mécanique. L'algorithme que nous proposons est relativement rapide et facile à mettre en oeuvre. Nous comparons les résultats obtenus sur des images synthétiques et satellitaires avec ceux produits par un modèle stochastique avec régularisation de Potts.

Mots-clés : Modèle variationnel, classification, théorie de transitions de phases, régularisation avec préservation des contours, minimisation, images satellitaires.

Contents

1 Introduction

Variational approach and Partial Differential Equation (PDE) models have shown to be efficient for a wide variety of image processing problems such as restoration and edge detection [2, 6, 11, 17, 34, 35, 37, 38, 46], or shape segmentation with active contours [16, 28, 32]. Nevertheless, the notion of classification, which consists of assigning a label to each site of an image, has rarely been introduced in a variational formulation (continuous models) mainly because the notion of class has a discrete nature. Many classification models can be found in the field of stochastic approaches (discrete models), with the use of Markov Random Field (MRF) theory as for instance in [10, 12, 18, 19, 29, 31, 33]. Structural approaches as splitting, merging and region growing models [40], and few models as a combination of statistical and deterministic technics [47, 49] have also been developed for image classification. But, to our knowledge, very few works have been conducted in the field of classification by the use of variational models.

Most of the images we are dealing with need to be restored because of the presence of noise. A lot of variational models have been developed for restoration with edge detection. In many applications, as satellite imaging, the goal is to obtain a region segmentation in addition to restoration. The principle is based on partitioning the image in different areas (i.e. in different classes), each area being characterized by a feature. The feature criterion we are interested in is the spatial distribution of intensity. Of course, other discriminant features than intensity can be used by considering suitable parameters. Within this framework, a region will be defined as the set of pixels belonging to the same class, and a class is characterized by parameters linked to the spatial distribution of intensity for instance (mean and standard deviation if we make a Gaussian assumption). Hence, we introduce a new functional for classification including restoration with edge regularization. The final results are the image of classes , the restored image and the map of regularized dicontinuities. The solution will be made of homogeneous regions, or classes, separated by sharp regularized boundaries.

In mechanics, many authors have studied the stability of systems containing several instable components through the Van der Waals-Cahn-Hilliard theory of phase transitions [1, 7, 8, 9, 13, 14, 15, 20, 36, 39, 41, 42, 44, 45]. These components may be liquids having different levels of density distribution. The stable configuration is proven to be compound of homogeneous regions separeted by sharp interfaces having minimal length. We propose to apply these results to a variational model for image classification.

First, the problem statement is considered. We then recall the Van der Waals-Cahn-

Hilliard theory of phase transitions introduced in mechanics. We also examine the relationship between this model and image classification, stressing the similarities of this theory with a region segmentation including a restoration process.
In the next section, we propose a new functional devoted to image classification in addition to an anisotropic smoothing capability. By introducing a new term related to the region segmentation, we impose a level constraint, each level being related to a specific class. We also discuss about the behaviour of this term. The algorithm we use to minimize this functional is based on the half-quadratic method presented in [17, 22, 46].
In the last section, we test the proposed model on synthetic and satellite images. These results are examined and compared with the ones obtained with a stochastic approach based on a Potts model.

2 Problem statement

2.1 Notation

The notation used hereafter are the following:
General notation:

- Ω is an open bounded subset of $\mathcal{R}^n$,
- u and f are elements of $L^1(\Omega, \mathcal{R})$,
- $\nabla = (\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n})^t$ is the gradient operator,
- $div(f) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}$ is the divergence operator applied to f.
- The Total Variation (TV) of f is defined as:

  $$TV(f) = \int_\Omega |Df| = sup\{\int_\Omega f.div(g);\ g \in C_0^1(\Omega, \mathcal{R}^n)\ and\ |g(x)| \leq 1\}.$$

  If $f \in C^1(\Omega)$ then $TV(f) = \int_\Omega |\nabla f|$.
- $BV(\Omega)$ is the space of functions with Bounded total Variation (BV) [24]:

  $$BV(\Omega) = \{f \text{ such that } TV(f) < +\infty\}.$$

  This space contains the functions which can be discontinuous through curves, and therefore this space is suitable for the study of free discontinuity variational problems as the segmentation one.

- The perimeter of the set $A \subset \Omega$ is defined as $Per_\Omega(A) = TV(\chi_A)$ where χ_A is the characteristic function of A.
- $H(A)$ is the $n-1$ dimensional Hausdorff measure of A, and if the boundary ∂A of A is smooth we have:

$$H(\partial A \cap \Omega) = Per_\Omega(A) = TV(\chi_A).$$

Classification notation:

- $S = \{s\}_{s=1...N}$ the set of sites,
- L_s represents the label attributed to the site s,
- Λ the set of admissible labels, $card\Lambda = M$ and $L_s \in \Lambda$,
- C_i is the i^{th} class characterized by a label in Λ.

Remarks:

- We will sometimes use f instead of $f(x)$ when no confusion is possible.
- We use the terminology "site" or "pixel", denoting a point of the image, even in the continuous case in order to make a connection between the variational classification (partitioning in regions R_i) and discrete stochastic labelling (partitioning in classes C_i). All the elements of a region R_i belong to the same class.
- We will either say "classify a site" or "attribute a label to this site".

2.2 Problem statement

Most of the images we are dealing with are corrupted by different sources of noise. The collected data contain false informations that must be removed or softened. Let assume that the noise $\eta : \mathcal{R}^n \to \mathcal{R}$, ($n = 2$ or 3), has a white gaussian distribution of intensity. If η is an additive noise, the linear degradation model is of the form:

$$o(x) = f(x) + \eta(x), \quad \forall x \in \Omega, \tag{1}$$

where o denotes the collected data. The goal of the restoration process is to retrieve f from o through the minimization of the functional $J(f) = \int_\Omega \big(f(x) - o(x)\big)^2 dx$ for instance. This problem is generally ill-posed (in the sense of Hadamard [26]),

and the regularization theory leads to a well-posed formulation. The basic idea of a regularization method consists of imposing some constraints on the solution f, and so the space of admissible solutions is restricted. For instance, to obtain smooth solutions, a Thikonov regularization term $\int_\Omega |\nabla f(x)|^2 dx$ can be introduced in the functional J [48].
Edges are probably the most important low level features in image processing, and are usually defined as a sharp transition of intensity level. Therefore, the restoration methods must be able to remove noise while preserving thin structures. Tikhonov regularization induces an isotropic smoothing, and edges are not preserved. In the field of variational approaches, many authors have worked on anisotropic smoothing as in [6, 11, 17, 34, 35, 37, 38, 46] for example. The main idea is to define a functional whose minimization provides a restored solution with preserved edges. Anisotropic smoothing must be able to remove noise without smoothing the edges. A common anisotropic smoothing model is based on the minimization of the following functional [6, 11, 17, 37]:

$$J(f) = \int_\Omega \big(f(x) - o(x)\big)^2 dx + \lambda^2 \int_\Omega \varphi(|\nabla f(x)|) dx. \tag{2}$$

The φ function, $\varphi : \mathcal{R} \to \mathcal{R}^+$, has at least the two following properties (see Fig. 1): if $t \in \mathcal{R}$ is close to zero, $\varphi(t)$ is quadratic or nearly quadratic (*smoothing* effect), and for large values of t, $\varphi(t)$ is linear or sub-linear (*edge-preserving* effect). We also have the basic asumptions on φ: $\forall t \in \mathcal{R}$, $\varphi(t) \geq 0$; $\varphi(0) = 0$; $\varphi(t) = \varphi(-t)$; φ continuously differentiable and $\forall t \geq 0$, $\varphi'(t) \geq 0$. The φ functions are usually classified in two categories, the convex ones and the nonconvex ones. When φ is convex, the theorical study about the minimization of J leads to results such as the existence of minimizers. On the other hand, for a nonconvex function, the theorical study is much more difficult (non uniqueness of the minimum, if it exists). Nevertheless, nonconvex functions are often used because they usually provide better results. Table 2 presents different φ functions commonly used.

The Mumford-Shah functional [38] is known as the most synthetic criterion for the segmentation process. We can find many publications about the mathematical soundness of this model, which belongs to the wide variety of free discontinuity problems, as well as many applications [3, 4, 6, 34, 35, 37, 46]. In addition to restoration, the notion of segmentation, i.e. getting an image partition, is introduced. The expression of the Mumford-Shah functional is:

$$J_{MS}(f, B) = \int_\Omega \big(f(x) - o(x)\big)^2 dx + \lambda^2 \int_{\Omega \setminus B} |\nabla f(x)|^2 dx + H(B), \quad f \in BV(\Omega) \tag{3}$$

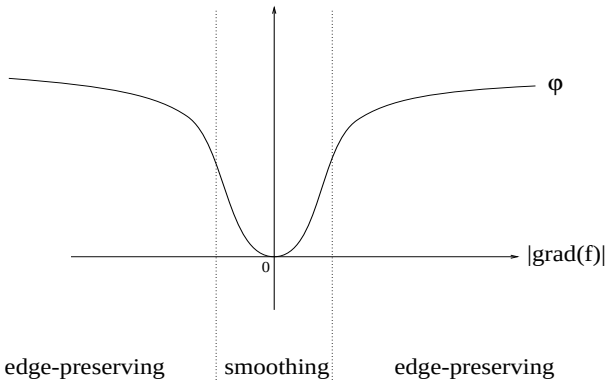


Figure 1: A typical behaviour of a φ function.

with B denoting the set of boundaries (unknown free discontinuities). The first term in (3) aims at keeping the solution close to the data. The second term, weighted by the regularization parameter λ, provides a smoothing process on homogeneous areas. The last term leads to boundaries having minimal length. So, the expected solution is compound of homogeneous areas separated by sharp regularized edges: a piecewise constant solution. The minimization of J_{MS} is a difficult task because the set B is unknown. In the field of free discontinuity problems, theorems of Γ-convergence can be used to approximate J_{MS} with a sequence of elliptic functionals, as for instance in [3, 4, 34, 35, 46].

In addition to the restoration and segmentation problems, we want to define a variational model also devoted to classification. Herein, a classification process means a segmentation with a labelling, i.e. every elements of a partition belong to the same class. This model must lead to three simultaneous processes: denoising, edge preservation and classification (*region segmentation*). In other words, for each site s we must be able to determine if s is a noisy pixel, if s is an edge element and which class s belongs to? (i.e. which is the label $L_s \in \Lambda$ of s?).
For the sake of simplicity, we make the following assumptions:

Hypothesis (H)

- Each site is characterized by its intensity level (grey level).
- Each class $C_{i;i=1..M}$ has a Gaussian distribution of intensity $N(\mu_i, \sigma_i)$ where μ_i and σ_i are respectively the mean and the standard deviation of

the class C_i. We will use vector notation:

$$\mu = \{\mu_i\}_{i=1..M} \quad \text{and} \quad \sigma = \{\sigma_i\}_{i=1..M}.$$

- The number M and the parameter vectors μ and σ of the classes are known (i.e. *supervised* classification).

This model relies on the minimization of a nonconvex functional (or *energy*) whose general formulation is:

$$J(f) = \int_\Omega \big(f(x) - o(x)\big)^2 dx + \lambda^2 \int_\Omega \varphi(|\nabla f(x)|)dx + \eta^2 \int_\Omega W(f(x); \mu, \sigma)dx. \quad (4)$$

The first two terms provide a restoration with edge-preservation. The last term W in (4) is a potential inducing a classification constraint. It takes into account the intensity and the parameters of the classes for the classification of each "pixel" x. W has got as many minima than the number of classes and imposes a level constraint on the solution. The real parameters λ and η permit to adjust the weight of each term.
In the next section, we will see that this model is close to the phase transitions problem studied in mechanics.

3 Van der Waals-Cahn-Hilliard theory of phase transitions

The Van der Waals-Cahn-Hilliard theory of phase transitions has shown to be efficient for the study of mechanical systems made up of a material compound of instable phases. This material may be either solid or liquid. For the sake of clarity, we expose the theory concerning two phases, and an extension to the M phases case can be found for instance in [7]. Thus, the two instable phases may represent either a chemical alloy as $Iron - Aluminum$ [1], or the distribution levels of a fluid [36]. The common problem in these different applications is to characterize the stability of such a system. This theory leads to the description of the stable configurations and to the analysis of the interface between the two phases while the system reaches its stability.
First, the Van der Waals-Cahn-Hilliard theory is briefly described. Then, we will stress that if we consider the image intensity as a material, this mechanical theory remains useful for image segmentation.

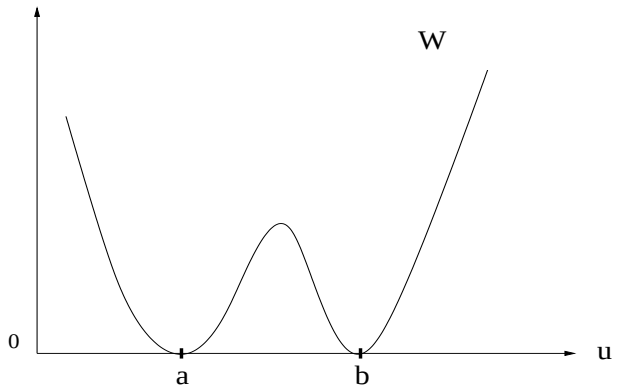


Figure 2: Example of double-well potential W.

3.1 Related works

Consider a dynamical system made of a single fluid whose Gibbs free energy, per unit volume, is a function W of the density distribution u. W has two minima, and is known as a *double-well potential* (see Fig. 2). This fluid is supposed to be under isothermal conditions and contained in a bounded domain Ω. The stable configurations of this system are obtained by solving the following variational problem P_ϵ (as for example in [7, 9, 20, 36, 44, 45]):

$$P_\epsilon \begin{cases} \inf_u E_\epsilon(u), \\ E_\epsilon(u) = \int_\Omega \Big[\epsilon|\nabla u(x)|^2 + \frac{1}{\epsilon} W(u(x))\Big] dx, \\ \text{with } u \text{ subject to the constraint: } \int_\Omega u(x)dx = m, \end{cases} \tag{5}$$

where Ω is an open bounded subset of $\mathcal{R}^n$ with Lipschitz continuous boundaries, and m is the total mass of the fluid. $|\nabla u(x)|$ represents the gradient modulus of $u(x)$. $W : \mathcal{R} \rightarrow \mathcal{R}^+$ is a continuous non negative function with exactly two minima a and b ($a < b$) such that $W(a) = W(b) = 0$. Furthermore, W is quadratic around a and b, and is growing at least linearly at infinity, see for instance [20] for explicit conditions on W (This study also holds for $W : \mathcal{R}^p \rightarrow \mathcal{R}^+$, as it is exposed for instance in [7, 20, 44]). The Van der Waals-Cahn-Hilliard theory has led to introduce the perturbation term $\epsilon|\nabla u|^2$, with ϵ small, which contributes to solve the failure concerning the uniqueness of the solution u of the problem: $\inf_u \int_\Omega W(u(x))dx$ with u such that $\int_\Omega u(x)dx = m$.

Concerning the study of the problem P_ϵ and its solutions as $\epsilon \to 0^+$, authors of [9, 20, 36, 44] used the Γ-convergence theory developed by De Giorgi [23]. If Ω and W verify the previously prescribed conditions, the main results are:

- E_ϵ Γ-converges to E_0 with:

$$E_0(u) = \begin{cases} K Per_\Omega(R_1) & \text{if } u(x) \in \{a; b\} \text{ a.e.} \\ +\infty & \text{otherwise,} \end{cases}$$

  with $R_1 = \{x \in \Omega / u(x) = a\}$ and K defined as:

$$K = 2 \inf_g \Big\{ \int_{-1}^{1} \sqrt{W(g(s))} |g^{'}(s)| ds; \;\; g \text{ piecewise } C^1,\, g(-1) = a,\, g(1) = b \Big\}.$$

- If u_ϵ is a sequence of minimizers of E_ϵ such that u_ϵ converges to u_0 in $L^1(\Omega)$, then u_0 is a solution of the problem:

$$\inf_u Per_\Omega(R_1) \quad \text{with} \begin{cases} u \in BV(\Omega), \\ W(u(x)) = 0 \;\; a.e., \\ \int_\Omega u(x) dx = m \end{cases}$$

- Any sequence (v_ϵ) such that $E_\epsilon(v_\epsilon) \leq constant < \infty \; \forall \epsilon$, admits a subsequence converging in $L^1(\Omega)$ (compactness in $L^1(\Omega)$).

Note that the reaction-diffusion problem related to P_ϵ is based on the following PDE (Ginzburg-Landau equation):

$$\frac{\partial \rho}{\partial t} = \epsilon \Delta \rho - \frac{1}{\epsilon} W^{'}(\rho), \quad \text{where } t \text{ is the evolving time parameter.} \tag{6}$$

Many authors have studied this PDE, as in [8, 14, 42], and have shown that the solution $\rho_\epsilon(x, t)$ of (6) as $\epsilon \to 0$ is piecewise constant with two admissible values a and b. Moreover, the evolution of the interface between the two regions $\{x \in \Omega / \rho_\epsilon(x, t) = a\}$ and $\{x \in \Omega / \rho_\epsilon(x, t) = b\}$ is a *motion by mean curvature*-type [1, 8, 30, 41, 42], inducing a certain regularity of the interface.

The double-well potential W can be interpreted as a level constraint term, forcing u to take one of the two values a or b. This constraint has an increasing weight as $\epsilon \to 0$. The perturbation term $\epsilon |\nabla u|^2$ has a regularization effect on the solution u

by avoiding the formation of singularities and by restricting the space of solutions. This perturbation term vanishes as $\epsilon \to 0$. This study is presented herein in the double-well case, even if several authors have worked on the multiple-well case and multiple fluids [5, 7, 13, 41, 45].

3.2 Analogy with image classification

We can transpose these previous works to image processing for a region segmentation and restoration process. Let $\Omega \subset \mathcal{R}^{2or3}$ be an open bounded domain, and $f : \Omega \to \mathcal{R}$ a function that represents for instance the intensity of each site. Consider a feature criterion of classification, only based upon the distribution of intensity, and let the image be compound of two regions $R_1 = \{x \in \Omega / f(x) = a\}$ and $R_2 = \{x \in \Omega / f(x) = b\}$. Let $o : \Omega \to \mathcal{R}$ denote the observed data, corrupted by an additive white Gaussian noise η, such that the linear degradation model is of the form $o(x) = f(x) + \eta(x),\ \forall x \in \Omega$. Let $\tilde{P}_\epsilon$ be the following system:

$$\tilde{P}_\epsilon \begin{cases} \min_f \tilde{E}_\epsilon(f), \\ \tilde{E}_\epsilon(f) = \int_\Omega \Big[\epsilon |\nabla f(x)|^2 + \frac{1}{\epsilon} W(f(x))\Big] dx, \\ \text{with } f \text{ subject to the constraint:} \quad \int_\Omega (f(x) - o(x))^2 dx \leq \sigma^2 \end{cases} \tag{7}$$

where σ is the standard deviation of the white Gaussian noise η, and W is as previously described (see subsection 3.1). The constraint in (7) imposes the solution f to remain close to the observed data.
According to the previous results, as $\epsilon \to 0$ the solution f of $\tilde{P}_\epsilon$ converges to a smooth segmented image whose sites only belong to R_1 or R_2 (a.e.), these two regions being separeted by sharp regularized edges (the smallest ϵ, the sharpest the transitions), and the set of boundaries having a minimal perimeter (see Fig. 3).

When ϵ is fixed, let examine the functional $\tilde{E}_\epsilon(f)$ to give an interpretation of the influence of each term on the solution of $\min_f \tilde{E}_\epsilon(f)$.

$$\tilde{E}_\epsilon(f) = \int_\Omega \Big[\underbrace{\epsilon |\nabla f|^2}_{\text{regularization term}} + \underbrace{\frac{1}{\epsilon} W(f)}_{\text{level constraint}} \Big] dx \tag{8}$$

The first term in (8) prevents the solution from the apparition of high level discontinuities. In the field of image restoration, this term induces a standard Tikhonov regularization (cf. section 2.2 and [48]). This regularization produces an isotropic

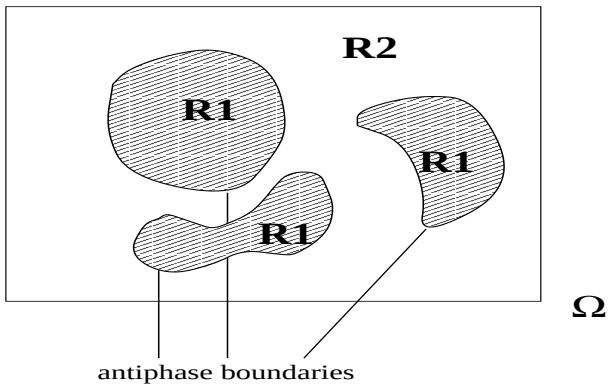


Figure 3: Segmented image with two regions R1 and R2.

smoothing. Nevertheless, the second term forces the solution to take only the two values characterizing the regions R_1 and R_2, these two values, a and b, are the labels of R_1 and R_2. Hence, even if the regularization is strong, the level constraint W should permit all the sites to stay sufficiently close to the levels a and b.
The convergence on ϵ induces an evolution of the weight of each term. The value of ϵ is initially high ("high" in the sense "non negligeable"). Therefore, the regularization process is firstly preponderant. As $\epsilon \to 0^+$, the influence of the level constraint is increasing while the smoothing term is vanishing. Roughly speaking, the convergence on ϵ permits in the first steps to remove the noise and then, when the image is sufficiently restored, we can start the classification with enough confidence.
Moreover, from the reaction-diffusion problem corresponding to (6), we know that the interface between R_1 and R_2 has a *motion by mean curvature-type*. This reflects the geometrical regularization operated on the edges.
The constraint in (7) measures the discrepancy between the solution and the data. So, the final image will be piecewise constant with regularized boundary (having a minimal perimeter), and will take into account the accordance with the data set.

4 Classification functional

4.1 Proposed functional

Based on the previous sections, we introduce the perturbation parameter ϵ in (4) by defining:

$$J_\epsilon(f) = \underbrace{\int_\Omega \big(f(x) - o(x)\big)^2 dx}_{\text{data term}} + \epsilon\lambda^2 \underbrace{\int_\Omega \varphi(|\nabla f(x)|)dx}_{\text{restoration term}} + \frac{\eta^2}{\epsilon} \underbrace{\int_\Omega W(f(x);\mu,\sigma)dx}_{\text{level constraint term}}. \quad (9)$$

The corresponding problem we are interested in is to find a solution $\tilde{f}$ such that

$$\tilde{f} = \lim_{\epsilon \to 0^+} \Big[\arg\min_f J_\epsilon(f)\Big]. \quad (10)$$

The first two terms of (9), if ϵ is fixed, are the ones of the restoration with edge-preservation functional (2). The last term is a level constraint inspired by the Van der Waals-Cahn-Hilliard theory of phase transitions, such that W attracts the values of $f(x)$ towards the labels of the classes. In our approach, since a class C_i is characterized by its mean μ_i and standard deviation σ_i, we have chosen to design C_i by the label μ_i. Hence, the set of admissible labels Λ is the vector μ.
In accordance with the hypothesis (H) stated in section 2.2, the potential $W : \mathcal{R} \to \mathcal{R}^+$ must take into account the Gaussian distribution property of the classes. W has its M minima on the values μ_i such that $W(\mu_i) = 0,\ \forall\, i$. We have constructed a potential W such that $W \in C^1(\mathcal{R})$ and W is piecewise parabolic. From Fig. 4, around μ_i, W is of the form $W(x) = P_i(x) = \frac{w_i}{{\sigma_i}^2}(x - \mu_i)^2$, where w_i is a positive weighting constant. Then, we construct the quadratic piecewise $Q_i(x) = k_i - c_i(x - p_i)^2$ that joins P_i and P_{i+1} respectively at the points α_i and β_i (Q_i is a *parabolic junction*). The parameters α_i, β_i, k_i, c_i ($c_i > 0$) and p_i are computed from the expression of P_i and P_{i+1}. Q_i joins P_i and P_{i+1} ($C^1(\mathcal{R})$-junction) such that $[\alpha_i, \beta_i]$ is as small as possible to get a potential W whose parabolics P_i are "predominant" in comparison with the parabolic junctions. For each set of parameters (μ, σ), we get a different potential W. W is quadratic around its minima and is growing faster than linearly at infinity (expected properties for W defined in section 3.1).
We assume that φ is a regularization function (as the ones of Tab. 2) having the characteristics exposed in section 2.2 and shown on Fig. 1.
Let ϵ be fixed. A minimizer of J_ϵ, if it exists, is a solution f_ϵ satisfying the expected properties: restored, edge-preserved (if φ is different from a "Tikhonov" regularization function, see Tab. 2) and level-constrained. As $\epsilon \to 0$, during the first steps

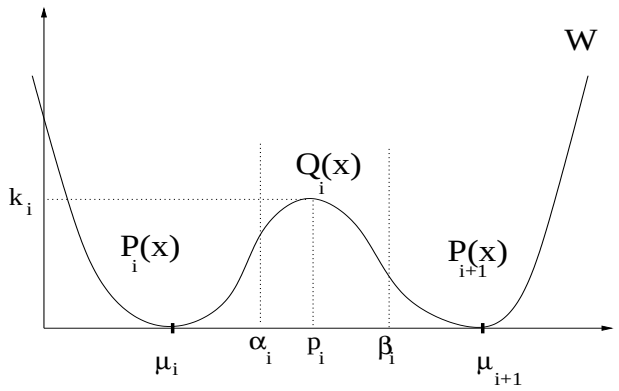


Figure 4: Construction of the piecewise parabolic potential W.

of the convergence the weight of the level constraint term is quite insignificant; only the restoration process occurs. As stressed in the previous section, as ϵ decreases, we progressively get a softened anisotropic diffusion while raising the classification process.
Note that the solution we are looking for is a continuous variable (as the intensity level for example) and not a discrete one as in a stochastic labelling [10, 18, 19, 29]. The classification is implicitly defined through the level constraint term in (9) which attracts the admissible values of the solution towards the M labels of the classes.

Remark: in the following, we will omit the convergence parameter ϵ indexing f, thus f_ϵ will be replaced by f, denoting the solution of the minimization of J_ϵ.

4.2 About the classification term W

Let summarize the properties of the level constraint potential W we expect, according to the assumptions (H) we made on the classes, and with respect to the Van der Waals-Cahn-Hilliard theory:

1. **Number of minima:** W takes into account the number and the parameters of the classes.

2. **Localization of the minima:** for M classes, W must have M minima on every site s such that $f(s)$ is close to μ_i. In other words, the label of a class C_i is the corresponding mean μ_i. According to section 3.1, a region R_i will be

defined such as $R_i = \{x \in \Omega / f(x) = \mu_i\}$ and each element of R_i belongs to C_i.

3. **Values of the minima:** W is constructed such that $W(\mu_i) = 0\ ,\ \forall\, i = 1..M$.

4. **Discrimination of the classes:** W should be able to discriminate two classes having respective distributions $N(\mu_i, \sigma_i)$ and $N(\mu_j, \sigma_j)$ if $\mu_i \neq \mu_j$ even if μ_i is close to μ_j and even in the case where σ_i and σ_j are such that the two distributions have a wide overlapping area (i.e. mixed distributions).

Points 1 to 3 are taken into account through the construction of W. There exists several ways of constructing a potential W satisfying points 1 to 3, but a piecewise quadratic $C^1(\mathcal{R})$ potential seems to be the easiest one, especially since the derivative W' is piecewise linear. When we have M classes, the expression of W is (see Fig. 4):

$$W(x) = \begin{cases} \frac{w_1}{{\sigma_1}^2}(x-\mu_1)^2 & \text{if } x \leq \alpha_1 \\ \frac{w_n}{{\sigma_n}^2}(x-\mu_n)^2 & \text{if } x \geq \beta_{n-1} \\ \frac{w_i}{{\sigma_i}^2}(x-\mu_i)^2 & \text{if } \beta_{i-1} \leq x \leq \alpha_i \\ k_i - c_i(x-p_i)^2 & \text{if } \alpha_i \leq x \leq \beta_i \end{cases} \tag{11}$$

Point 4 is the most delicate. We want to define a potential W having M wells (point 1) sufficiently marked to obtain M responses. This means that we want to avoid a configuration in which a well is hidden by the one of a neighbouring class. To illustrate such a phenomena and to explain how to overcome this difficulty, let assume that we deal with exactly 3 classes with respective parameters (mean; standard deviation): (12.0; 8.0), (13.0; 6.0) and (16.0; 4.0). Let examine the influence of the *local* parameters w_i weigthing each quadratic P_i of W (see (11)). The widening of the local parabolic P_i is determined by the ratio $\frac{w_i}{{\sigma_i}^2}$. The value of σ_i is fixed, and the value of w_i is selected by the operator. So, a variation of w_i induces a variation of the width of the well formed by the parabolic P_i such that the most the value of w_i, the narrowest the well. Fig. 5 presents the 3-well potential W for which $w_1 = w_2 = w_3 = 1.0$. The classes whose labels (mean values) are 12.0 and 13.0 are nearly not distinguishable. By setting $w_1 = w_2 = 10.0$ and $w_3 = 1.0$ we make the two wells emerge, and we can see on Fig. 6 that the two neighbouring classes are now distinguishable.

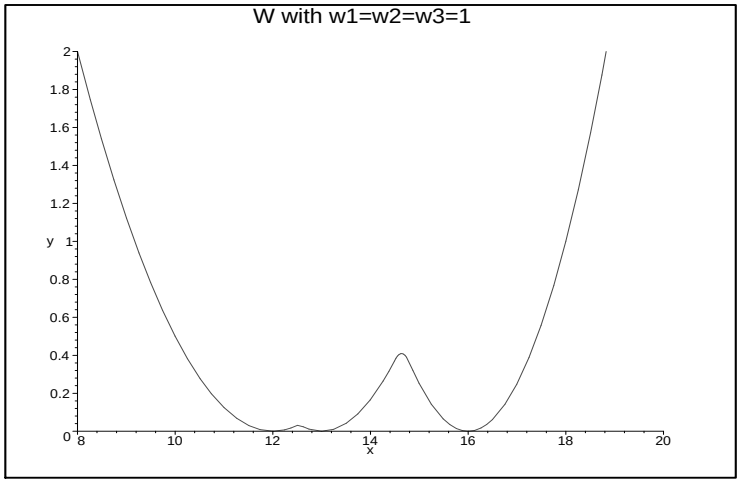


Figure 5: Potential W for 3 classes (see section 4.2). The weighting parameters w_i of P_i (see (11)) are $w_1 = w_2 = w_3 = 1.0$.

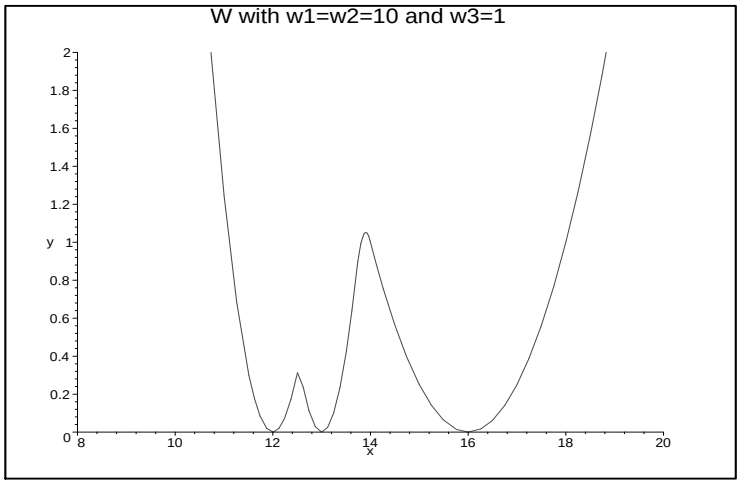


Figure 6: Potential W for 3 classes (see section 4.2). The weighting parameters w_i of P_i (see (11)) are $w_1 = w_2 = 10.0$ and $w_3 = 1.0$.

	MECHANICS	→	IMAGE PROCESSING
(a)	constraint in the problem (5)	→	penalty introduced in the functional (9)
(b)	potential with 2 wells	→	potential with M wells (classes)
(c)	perturbation term $\epsilon\|\nabla u\|^2$	→	regularization term $\epsilon\varphi(\|\nabla f\|)$

Table 1: From the phase transitions theory to the proposed classification functional

4.3 Soundness of the proposed model

Table 1 summarizes the transpositions we made from mechanics (5), for which we have theorical results at our disposal, to image classification through the mini-minization of (9). We propose to examine extrapolations (a) to (c) in Table 1 by distinguishing the ones which are mathematically justified from the heuristic ones.

(a) - from constraint to penalty -

In [7, 9, 36, 44], the mathematical results about the theory of phase transitions rely on the Γ-convergence theory. One of the main advantages of using this kind of variational convergence rather than only convergence of minimizers, is that the results remain unchanged with a continuous perturbation on the functional [23]. In particular:

$$\begin{array}{ll} \text{if} & F_\epsilon(u) \xrightarrow{\Gamma} F_0(u) \quad \text{as } \epsilon \to 0^+ \\ \text{then} & F_\epsilon(u) + \int_\Omega \big(u(x) - o(x)\big)^2 dx \xrightarrow{\Gamma} F_0(u) + \int_\Omega \big(u(x) - o(x)\big)^2 dx \quad \text{as } \epsilon \to 0^+ \end{array}$$

with $o : \mathcal{R}^n \to \mathcal{R}$ a function in $L^2(\mathcal{R})$, denoting the data. In addition to the characteristics of minimizers presented in section 3.1, minimizers f_ϵ of (9) will take into account the observed data o. Hence there will be a trade-off between the regularity of the solution and its level of accordance with the data.

(b) - from 2 wells to M wells -

Most of authors have studied the double-well case [1, 8, 14, 36, 42, 44]. Some of

them have treated the 3-well case as in [41, 45], and the general case corresponding to M wells is presented by Baldo in [7]. Baldo has extended the previous results to the case of $u : \mathcal{R}^n \rightarrow \mathcal{R}^p$ and for a potential $W : \mathcal{R}^p \rightarrow \mathcal{R}_+$ having M zero-valued minima.

(c) - from quadratic perturbation to regularization using a φ function -

Let note $\Theta : \Omega \times \mathcal{R} \times \mathcal{R}^n \rightarrow \mathcal{R}^+$ such that (see (5)):

$$\Theta(x, u, \epsilon \nabla u) = E_\epsilon(u).$$

Owen and Sternberg [39] have proven that for any convex Θ function of $\epsilon \nabla u$ (other weaker conditions on Θ are required in their article), u_ϵ, the solution of P_ϵ as $\epsilon \rightarrow 0$, still has the same characteristics than the ones previously exposed for an isotropic perturbation $\epsilon^2 |\nabla u|^2$ (the rescaled perturbation of (5)). However, Θ must have (at least) a parabolic behaviour with respect to $\epsilon \nabla u$ at infinity. Thus, if φ is the Tikhonov regularization function (see Table 2) then $\epsilon^2 \varphi(|\nabla f|) = \epsilon^2 |\nabla f|^2$ and the theory holds. If φ is another regularization function, the theorical soundness of (9) is an open problem.

4.4 Minimization and algorithm

Let describe the strategy we have adopted to minimize (9) with respect to f.
If f is a minimizer of (9), then a necessary condition is that the derivative of the functional J_ϵ applied to f is equal to zero. Using the Euler-Lagrange equation to compute the derivative of J_ϵ, we obtain the following system:

$$\begin{cases} \big(f(x) - o(x)\big) + \frac{\eta^2}{2\epsilon} W'(f(x)) - \epsilon \lambda^2 div\big(\frac{\varphi'(|\nabla f(x)|)}{2|\nabla f(x)|} \nabla f(x)\big) = 0 \ \ \forall x \in \Omega, \\ \frac{\partial f(x)}{\partial n} = 0 \quad \text{on } \partial\Omega, \end{cases} \tag{12}$$

where n is the outward normal to the boundary domain $\partial\Omega$, and div denotes the divergence operator. Let remark the nonlinearity of the PDE in (12) although $W'(f(x))$ is piecewise linear with respect to $f(x)$ (see (11)). This nonlinearity is due to the fact that we do not have the Laplacian operator $\Delta f = div(\nabla f)$, which induces an isotropic smoothing (Tikhonov regularization), but an anisotropic smoothing operator such that the gradient in the divergence is weighted by the coefficient $\frac{\varphi'(|\nabla f|)}{2|\nabla f|}$.
To obtain an edge-preserving process, according to [17], we impose the following constraints on φ:

(a) $\lim_{t\to 0} \frac{\varphi'(t)}{2t} = \gamma > 0$ and $\gamma < \infty$,
(b) $\lim_{t\to \pm\infty} \frac{\varphi'(t)}{2t} = 0$,
(c) $\frac{\varphi'(t)}{2t}$ symetrical and strictly decreasing on $[0; +\infty[$.

Condition (a) is related to an isotropic smoothing on homogeneous areas, while (b) leads to preserve edges from excessive smoothing. Condition (c) permits to avoid instabilities.
Let ϵ be fixed. The approach we adopt to minimize J_ϵ with respect to f is based on the *half quadratic regularization* method [17, 22]. This method permits to simplify the minimization of J_ϵ, by avoiding the difficulties related to the non linearity of the PDE in (12). The basic idea is to introduce an auxiliary variable which leads to find minimizer of (9) through the resolution of linear PDE's. Let assume that φ is a function verifying conditions (a) to (c). In the precited references, it has been proven that φ can be written as:

$$\varphi(t) = \inf_b [bt^2 + \psi(b)] \tag{13}$$

where the function ψ is derived from φ: $\psi(b) = g((g')^{-1}(b)) - b(g')^{-1}(b)$ with $g(t) = \varphi(\sqrt{t})$ (see [17, 22] for more details), and ψ is strictly convex. b is an auxiliary variable defined in (13) and whose value is given by the formula: $b = \frac{\varphi'(t)}{2t}$. According to the half quadratic regularization, we get:

$$J_\epsilon(f) = \min_b J_\epsilon^*(f, b) \tag{14}$$

with

$$J_\epsilon{}^*(f,b) = \int_\Omega \big(f(x) - o(x)\big)^2 dx + \epsilon\lambda^2 \int_\Omega \Big[b(x)|\nabla f(x)|^2 + \psi(b(x))\Big] dx + \frac{\eta^2}{\epsilon} W(f(x)) dx. \tag{15}$$

$J_\epsilon^*(f, b)$, called the *augmented energy*, is convex with respect to the auxiliary variable b (with f fixed) and piecewise quadratic with respect to f (with b fixed). According to the conditions on φ, the values of b are scaled from 0 to γ, and in our case the φ function we use is such that $\gamma = 1$ (see Table 2). The variable b marks the location of dicontinuities such that $b(x) \sim 0$ if x belongs to an edge and $b(x) \sim 1$ on homogeneous areas. From (14), the minimization of J will be transposed into the minimization of J_ϵ^* with respect to (f, b). $J_\epsilon{}^*$ is piecewise quadratic in f and

	$\varphi(t)$	$\frac{\varphi^{'}(t)}{2t}$	convexity/conditions (a) to (c) satisfied
Tikhonov [48]	t^2	1	yes/no
Total Variation [43]	$\lvert t\rvert$	$\frac{1}{2\lvert t\rvert}$ (if $t \neq 0$)	yes/no
Geman & McClure [21]	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	no/yes
Hebert & Leahy [27]	$\log(1+t^2)$	$\frac{1}{1+t^2}$	no/yes
Green [25]	$\log(\cosh(t))$	$\frac{\tanh(t)}{2t}$ (if $t \neq 0$)	yes/yes
Hyper Surfaces [17]	$\sqrt{1+t^2}-1$	$\frac{1}{2\sqrt{1+t^2}}$	yes/yes

Table 2: Some edge-preserving φ functions.

convex in b. The auxiliary variable b is related to the presence of discontinuities, and $W^{'}(f(x))$ is piecewise linear (locally linear) with respect to $f(x)$. Let note

$$W^{'}_{local}(f(x)) \;=\; 2[A_{local}f(x) \;+\; B_{local}],$$

with A_{local} and $B_{local} \in \mathcal{R}$. From (11) we notice that the expression of $W^{'}_{local}$ (i.e. the values of A_{local} and B_{local}) depends on the value of $f(x)$.
By the use of Euler-Lagrange equation, the minimization of J^*_ϵ with respect to f leads to a (locally) linear PDE. The minimization of J_ϵ while $\epsilon \to 0^+$ is operated through the algorithm:

- $\hookrightarrow$ initialize f
- $\hookrightarrow$ initialize ϵ
  - $\hookrightarrow$ repeat
    - $\diamond$ $\min_b J^*(f,b)$ with f fixed:
      $$b(x) = \frac{\varphi^{'}(\lvert\nabla f(x)\rvert)}{2\lvert\nabla f(x)\rvert}$$
    - $\diamond$ $\min_f J^*(f,b)$ with b fixed, i.e. solve the PDE:
      $$f(x) - \epsilon\lambda^2 div\big[b(x)\nabla f(x)\big] + \frac{\eta^2}{\epsilon}A_{local}f(x) = o(x) - \frac{\eta^2}{\epsilon}B_{local}$$
  - $\hookrightarrow$ until convergence on f
- $\hookrightarrow$ decrease ϵ

The PDE resulting from $\min_f J^*(f,b)$ with b fixed is solved using a conjugate gradient algorithm.

Label attribution:
The highest the number of loops on the decreasing of ϵ, the closest the values of the

solution to the minima μ_i of W. For practical reasons (computational time) and from experimental qualitative results, we stop the convergence before ϵ reaches zero. Therefore, the values of the solution are not always strictly equal to (but very close to) the labels μ_i. In order to always get a perfect labellized image we make the following operation on the solution of the previous algorithm: for each site s of the solution, if $p_i \leq f(s) \leq p_{i+1}$ (see Fig. 4) the label of s is set to μ_i.

5 Experimental results and comparison

We present experimental results on both synthetic and satellite images. We also compare the results provided from the variational model with the ones obtained by using a stochastic model for classification. We first describe this stochastic model. Then, we compare the results we obtain by the minimization of (9) using the deterministic algorithm presented in section 4.4 to those obtained with the stochastic model.

5.1 Description of the stochastic model

The stochastic model used hereafter is based on the Markov Random Field (MRF) theory, and mentioned for instance in [10, 18, 19]. This model provides an image classification by assigning a label to each site. The main differences between the stochastic model and the variational one are: first a discrete framework instead of a continuous one, second the unknown is an image of labels and not the intensity, and finally the algorithm used to minimize the energy is stochastic.
Let S and Λ be defined as before. Let C be the set of cliques such that we define a first order MRF model (i.e. we will take into account a site and its four nearest neighbours). A global discrete labelling L consists of assigning a label $L_s \in \Lambda$ to each site $s \in S$. By the use of Bayes theorem, the estimated labelling $\hat{L}$ is obtained through the following rule:

$$\begin{aligned} \hat{L} = \arg\max_L P(L/O) &\overset{\text{Bayes}}{=} \arg\max_L P(O/L)P(L) \\ &= \arg\min_L(-\ln P(O/L)P(L)), \end{aligned} \tag{16}$$

with O denoting the observed data, and by omiting $P(O)$ which does not depend on L. $P(O/L)$ is supposed to be such that $P(O/L) = \prod_s P(O_s/L_s)$ (white invariant noise), where L_s is the label assigned to the site s, and with:

$$P(O_s/L_s) = \frac{1}{\sqrt{2\Pi}\sigma_{L_s}} \exp\Big[-\frac{(O_s - \mu_{L_s})^2}{2\sigma^2{}_{L_s}}\Big]. \tag{17}$$

The prior model is a Potts regularization such that:

$$P(L) = \exp\Big[-\frac{\beta}{T} \sum_{\{s,r\}\in C} \delta(L_s, L_r)\Big]$$
$$\text{with} \quad \delta(a,b) = \begin{cases} -1 \text{ if } a = b \\ +1 \text{ if } a \neq b \end{cases} \tag{18}$$

where T is the temperature parameter. In accordance with (16), the estimated labelling $\hat{L}$ is obtained by solving:

$$\hat{L} = \arg\min_L E(L), \quad \text{and } E \text{ is a global energy such that:}$$
$$E(L) = \frac{1}{T}\Big[\sum_{s=1}^{N}\Big(\ln\sqrt{2\Pi}\sigma_{L_s} + \frac{(O_s - \mu_{L_s})^2}{2\sigma^2{}_{L_s}}\Big) + \beta \sum_{\{s,r\}\in C} \delta(L_s, L_r)\Big]. \tag{19}$$

Using a MAP (Maximum a Posteriori) criterion, the optimization is made by simulated annealing with a Metropolis algorithm. In addition to the number and the parameters of the classes, the required input are: the initial temperature, the decreasing factor governing the temperature descent, and the value of β. The temperature may decrease slightly. The value of β depends on the level of homogeneity we expect. The minimization of (19) straightly leads to a labellized image, whereas in the variational framework we progressively (as $\epsilon \to 0$) get a labelling. In order to get classification results directly comparable to the ones provided from the variational model, we impose that the labels L_s take their values in $\{\mu_i\}_{i=1..M}$, μ_i being the mean of the classe C_i.

5.2 Experimental results

We have tested the proposed variational model on both synthetic and real satellite grey level images. The results on synthetic images are meaningfull to validate the model and to analyze the differences with the stochastic model described in section 5.1. All results were obtained on a 166 MHZ computer.

Remark: In the following "VC" will denote the variational classification and "SC"

the stochastic one.

In order to run the previous variational algorithm, we need input parameters that we experimentally adjust. We first assume that parameters μ, σ and the number M of classes are known (preliminary estimation or parameters given by an expert). In addition to the values of λ, η in (9) we introduce a practical parameter δ which aims at rescaling the gradient modulus: $\frac{|\nabla f|}{\delta}$. This parameter can be seen as a threshold fixing the value from which a discontinuity is assimilated to an edge. We then start from $\epsilon = 1$ and select its speed of decreasing. We initialize f to the data set o. From experiments, we have noticed that we need to increase the global regularization term (restoration and level constraint) i.e. the last two terms of (9) while $\epsilon \to 0$ in order to get better numerical stabilities and to avoid the reappearence of noise as the restoration term vanishes. When it is not mentioned, the selected φ function is the one proposed by Geman and Mc Clure (see Table 2 and [22]).

5.2.1 Synthetic images

The first noisy synthetic image considered is of size 128×128 pixels ("check" image) and contains four classes whose parameters are given in Table 3. The white Gaussian noise introduced is such that the SNR is 10 dB (variance ratio). The 4-well potential W in (9) is illustrated on Fig. 7. Fig. 8 presents the different output images we obtain from the minimization of (15): the restored image (SNR=27.9 dB), the map of discontinuities (image of auxiliary variable b: external boundaries) and the VC which is compared to the SC. If we compute the percentage of misclassified pixels, the VC leads to 0.23 % and the SC to 0.18%. This slight difference mainly comes from the small piece of white square overlapped by the triangle and the circle (see the classification results on Fig. 8). The computational time is 18 seconds for the VC and 78 seconds for the SC. Fig. 9 shows the SNR and number of misclassified pixels evolution while ϵ decreases. We can detect the step on ϵ from which the SNR and the number of errors are stable (this can determine the stopping value of ϵ). On Fig. 10 we show the influence of the kind of φ function on VC results (see Table 2). For a Tikhonov regularization, edges are oversmoothed. With a convex φ, there still are many misclassified pixels on the bounbaries. Best results are provided with the use of nonconvex φ functions, even if the theorical soundness of the model is an open problem.

class	mean μ_i	standard deviation σ_i
1 (black squares)	85.0	12.85
2 (triangle)	115.0	12.85
3 (disk)	145.0	12.85
4 (white squares)	175.0	12.85

Table 3: Parameters of the 4 classes for "check" image. The standard deviation is the same for each class, and is the one of the white Gaussian noise.

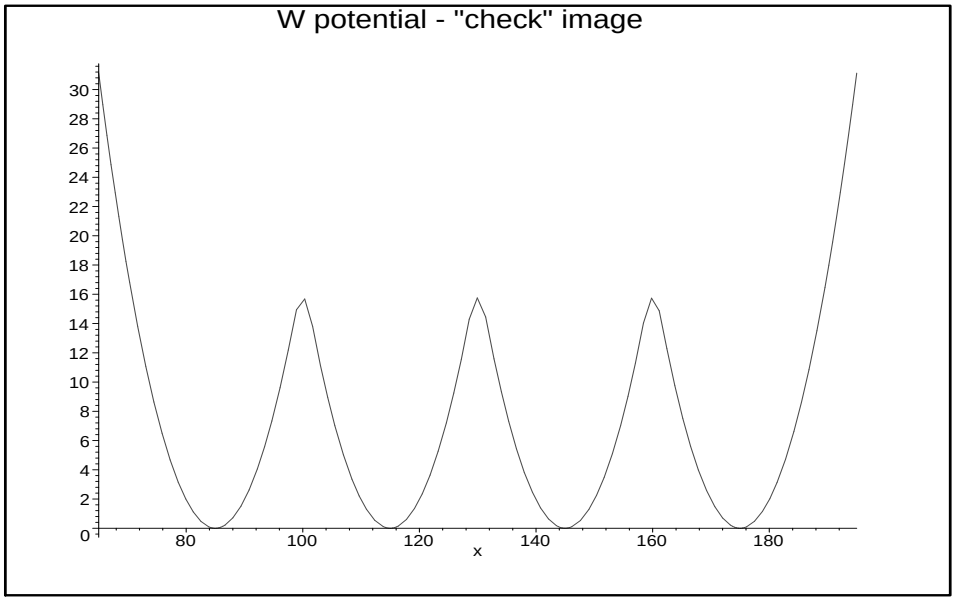


Figure 7: Potential W for "check" image. Since the standard deviation is the same for all the classes, the wells have the same width. The peaks seem to be nonsmooth, but this is due to the plot discretization.

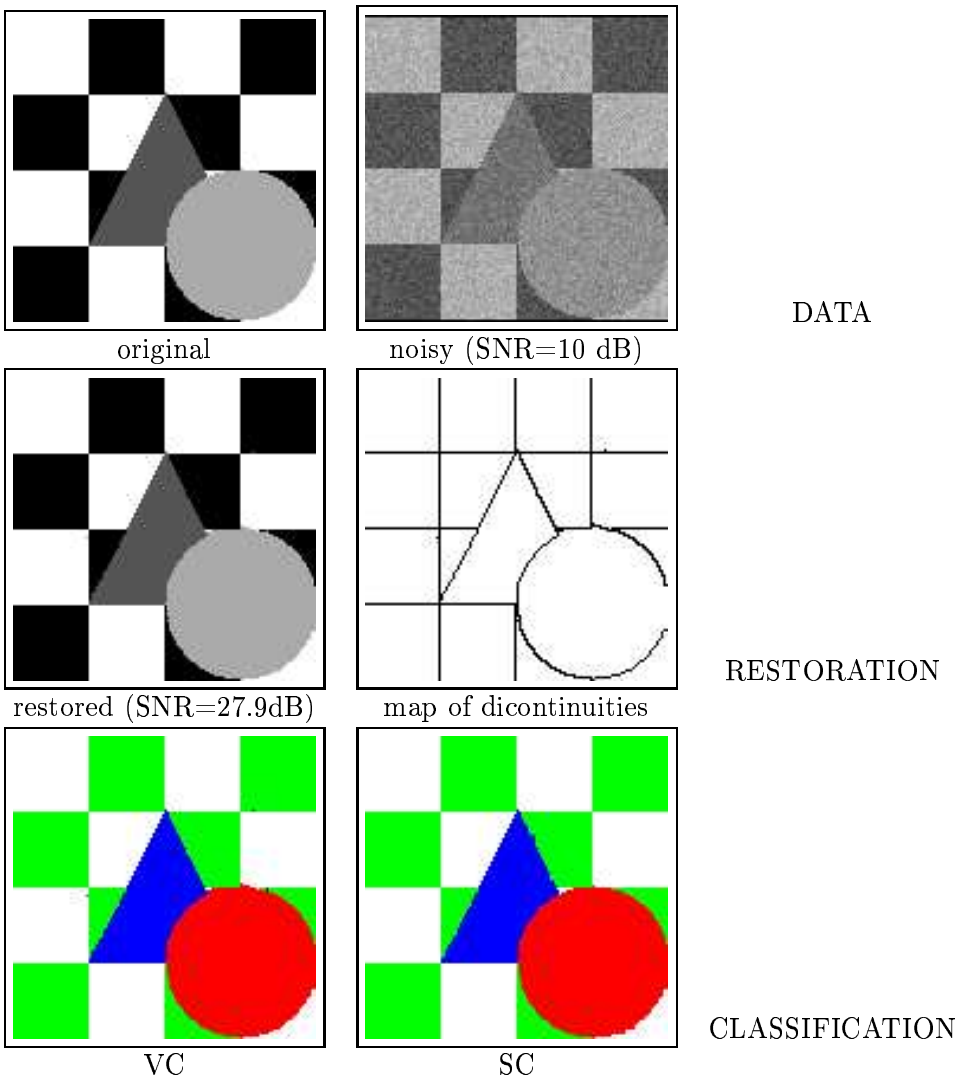


Figure 8: Results on "check" image. Variational parameters: $\lambda = 16$, $\delta = 15$, $\eta = 0.1$, 7 iterations on ϵ. Stochastic parameters: $\beta = 2.1$, 1000 iterations sor simulated annealing. Computational time: variational=18s., stochastic=78s.

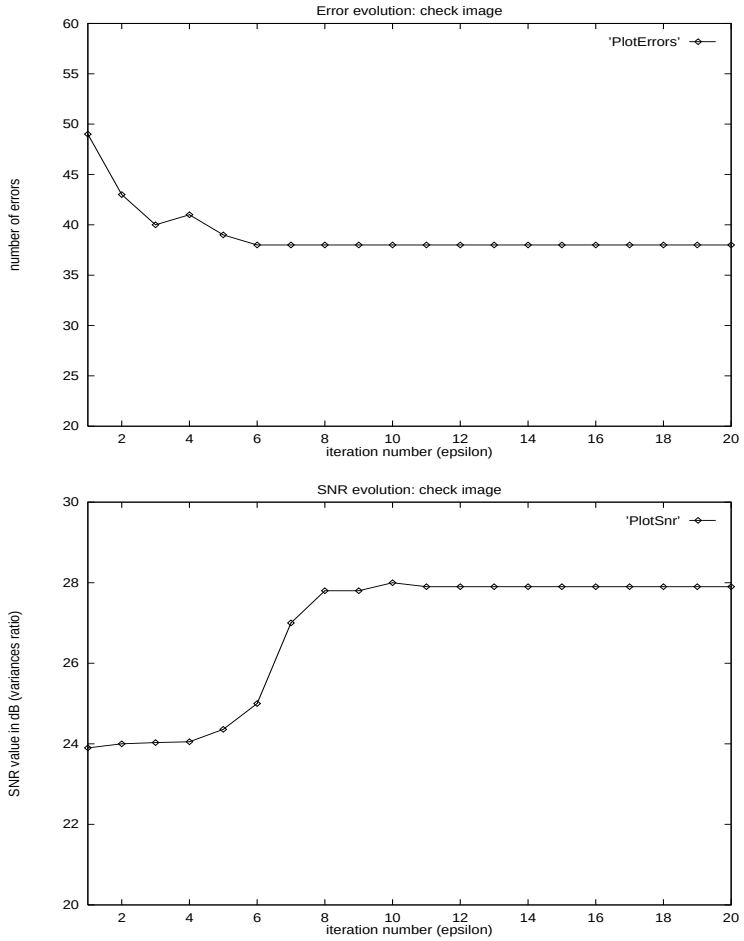
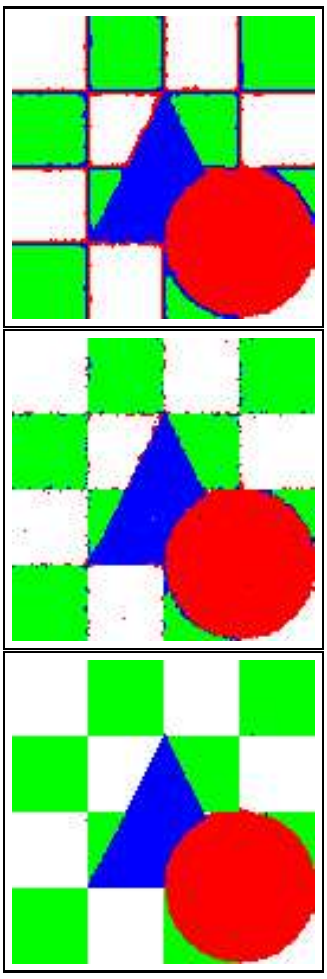


Figure 9: Number of misclassified pixels (top) and SNR value (bottom) with respect to the number of iterations on ϵ for "check" image.



$\varphi(t) = t^2$: convex (Tikhonov)

$\varphi(t) = \log(\cosh(t))$: convex (Green)

$\varphi(t) = \frac{t^2}{1+t^2}$: nonconvex (Geman & McClure)

Figure 10: Classification of "check" image with different φ function (see Tab. 2). Nonconvex functions provide better results even if there are no theorical proof concerning the solution in the nonconvex case. Convex functions lead to oversmooth results: we get damaged edges.

The second image is of size 256×256 pixels ("gdr" image), and the noisy version is such that the SNR is equal to 10 dB. This synthetic image comes from the database of the GdR ISIS, and was created to test segmentation and classification models. This image includes two main difficulties. The first one is the irregularity of the shape on bottom right handside. The second one is due to the grey level gradation on top right handside. We present results for two different sets of number and parameters of classes.

First, we have selected three classes whose parameters are summarized in Table 4. The corresponding 3-well potential is shown on Fig. 11; the wells are wide and clearly separated. On Fig. 12, we can see that the VC manages slightly better than the SC does the grey level gradation (cf. top right handside), and leads to less irregular boundary than the one of SC. Globally, VC results are close to the SC ones, but the computational time is 49 seconds for the VC and 285 seconds for the SC.

Then, we have chosen eight classes (see Table 5). From the 8-well potential shown on Fig. 13, we remark that classes number 6,7 and 8 are nearly undistinguishable. The classification results are presented on Fig. 14. We can notice that for this delicate configuration, the variational model is much more efficient (and faster) than the stochastic model. The SC is made of a lot of false regions, which is a consequence of the proximity of the classes. If we reduce β, which is the weight of Potts regularization, we get less false regions, but we get a lot of isolated misclassified pixels (due to the noise). The VC errors are mainly situated on boundaries of the irregular shape on bottom left, but this is the best trade-off: if we want to remove this misclassified pixels, we shall damage the edges. The computational time is 101 seconds for the VC and 671 seconds for the SC. On Fig. 15 we show the evolution of errors and SNR while ϵ decreases. Note the variation of the SNR evolution between iterations 30 and 40 before stability. This corresponds to the range for which the restored image becomes a piecewise constant image (the set of admissible values of $f(x)$ converges to the set of labels). Since the SNR is computed from the comparison with original image (grey levels), the SNR value at stability is not always the maximum. Especially herein, since we do not retrieve all the levels of the gradation of grey levels: in the stability domain beginning after iteration 30, f is piecewise constant with eight admissible values, whereas the original unnoisy image has 32 grey levels (see Fig. 16 and Fig. 17).

class	mean μ_i	standard deviation σ_i
1	100.0	5.56
2	128.0	5.56
3	160.0	5.56

Table 4: Parameters of the 3 classes for "gdr" image. The standard deviation is the same for each class, and is the one of the white Gaussian noise.

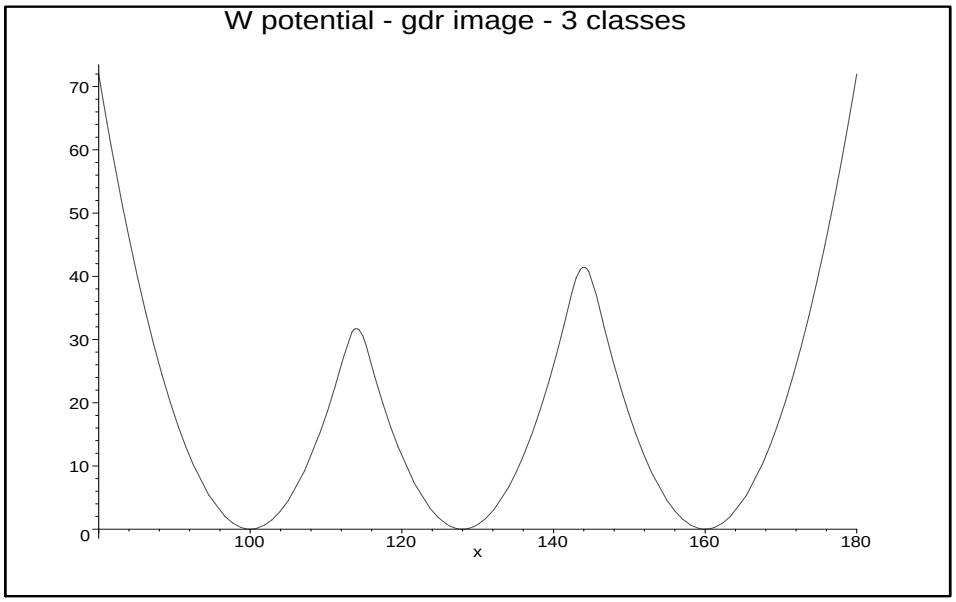


Figure 11: Potential W for "gdr" image with 3 classes. Since the standard deviation is the same for all the classes, the wells have the same width.

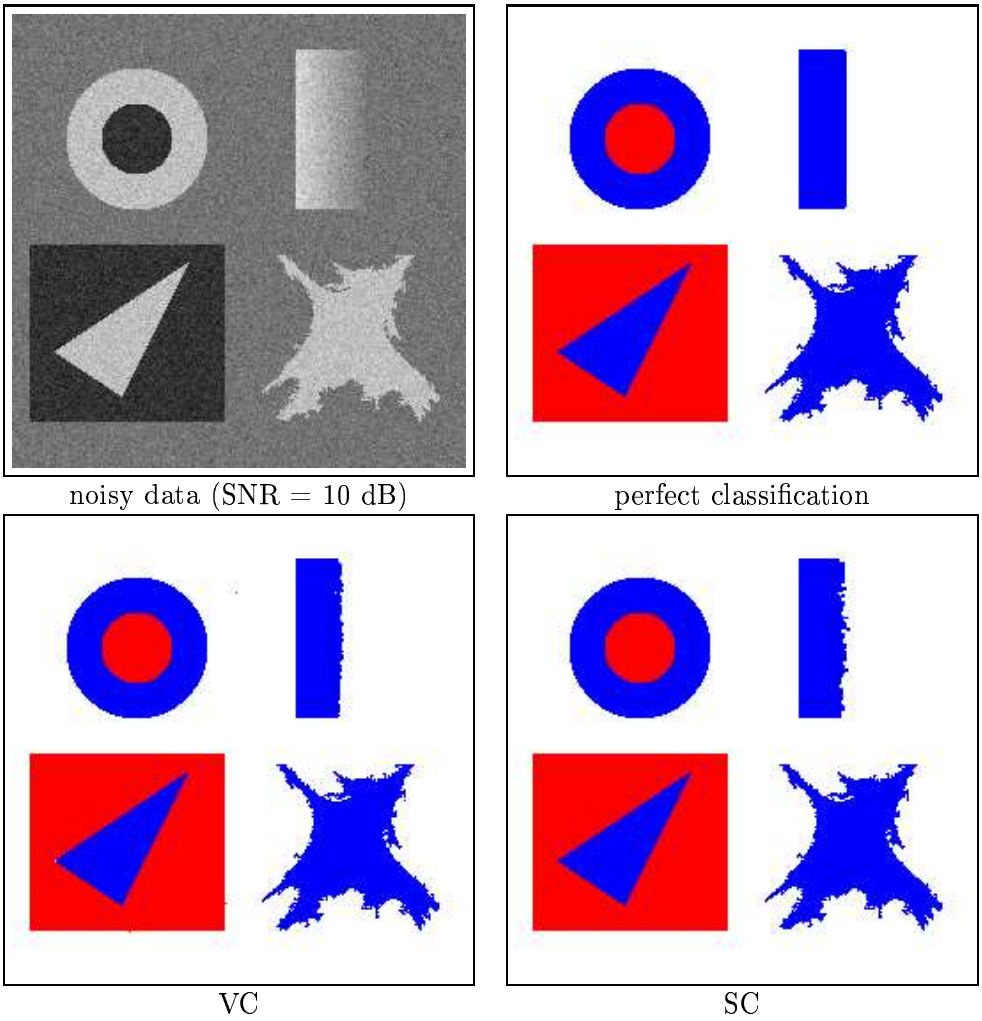


Figure 12: Results for a noisy synthetic "gdr" image containing 3 classes. Variational parameters: $\lambda = 5$, $\delta = 10$, $\eta = 0.05$, 16 iterations on ϵ. Stochastic parameters: $\beta = 4.0$, 1000 iterations for simulated annealing. Computational time: variational=49s., stochastic=285s.

class	mean μ_i	standard deviation σ_i
1	100.0	5.56
2	128.0	5.56
3	135.0	5.56
4	142.0	5.56
5	149.0	5.56
6	156.0	5.56
7	160.0	5.56
8	163.0	5.56

Table 5: Parameters of the 8 classes for "gdr" image. The standard deviation is the same for each class, and is the one of the white Gaussian noise.

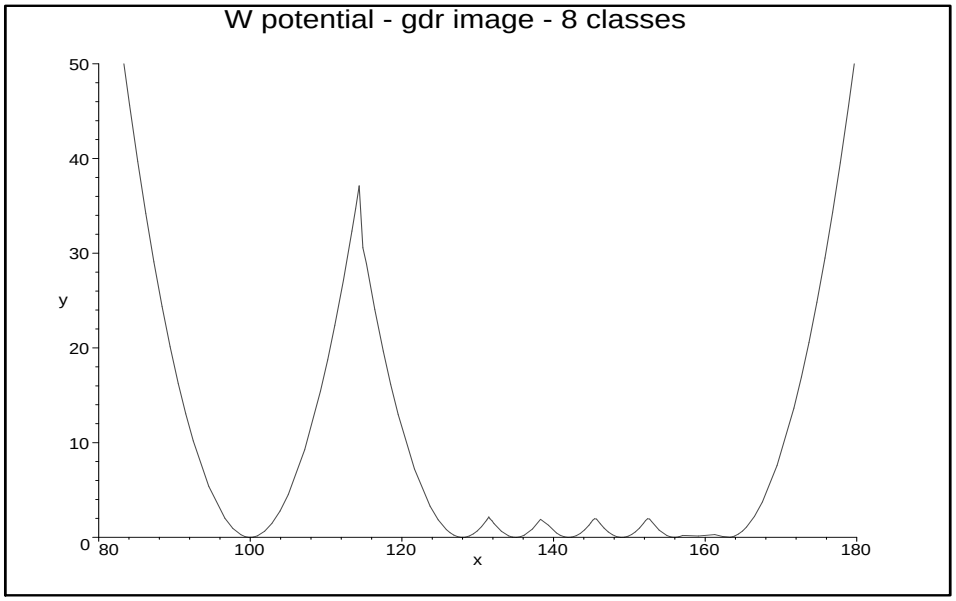


Figure 13: Potential W for "gdr" image with 8 classes. Let note that the classes whose mean values are 156.0, 160.0 and 163.0, are nearly undistinghuishable.

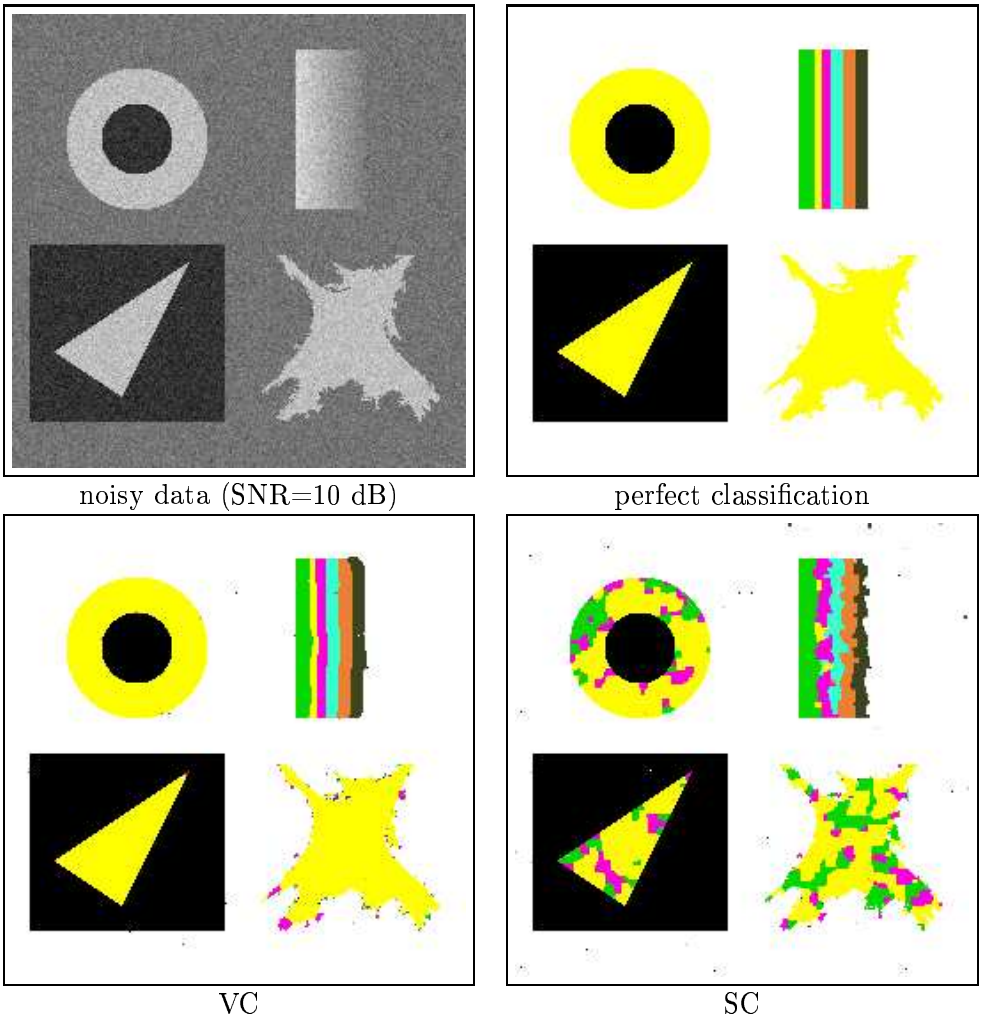


Figure 14: Results for a noisy synthetic "gdr" image containing 8 classes. Variational parameters: $\lambda = 60$, $\delta = 6$, $\eta = 0.05$, 35 iterations on ϵ . Stochastic parameters: $\beta = 1.3$, 2000 iterations for simulated annealing. Computational time: variational=101s., stochastic=671s.

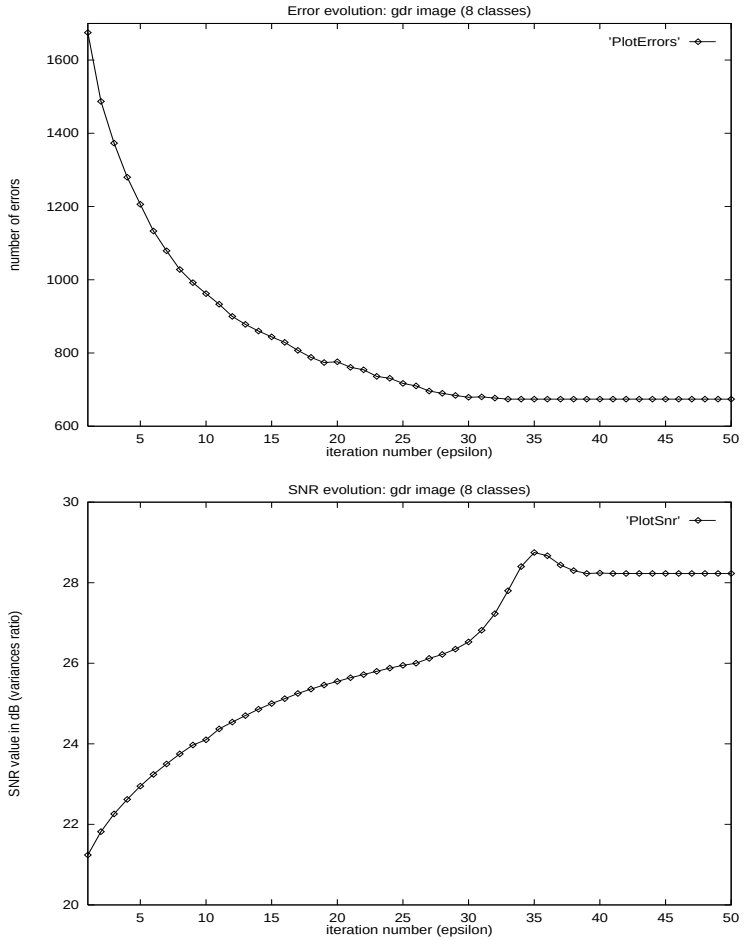


Figure 15: Number of misclassified pixels (top) and SNR value (bottom) with respect to the number of iterations on ϵ for "gdr" image with 8 classes.

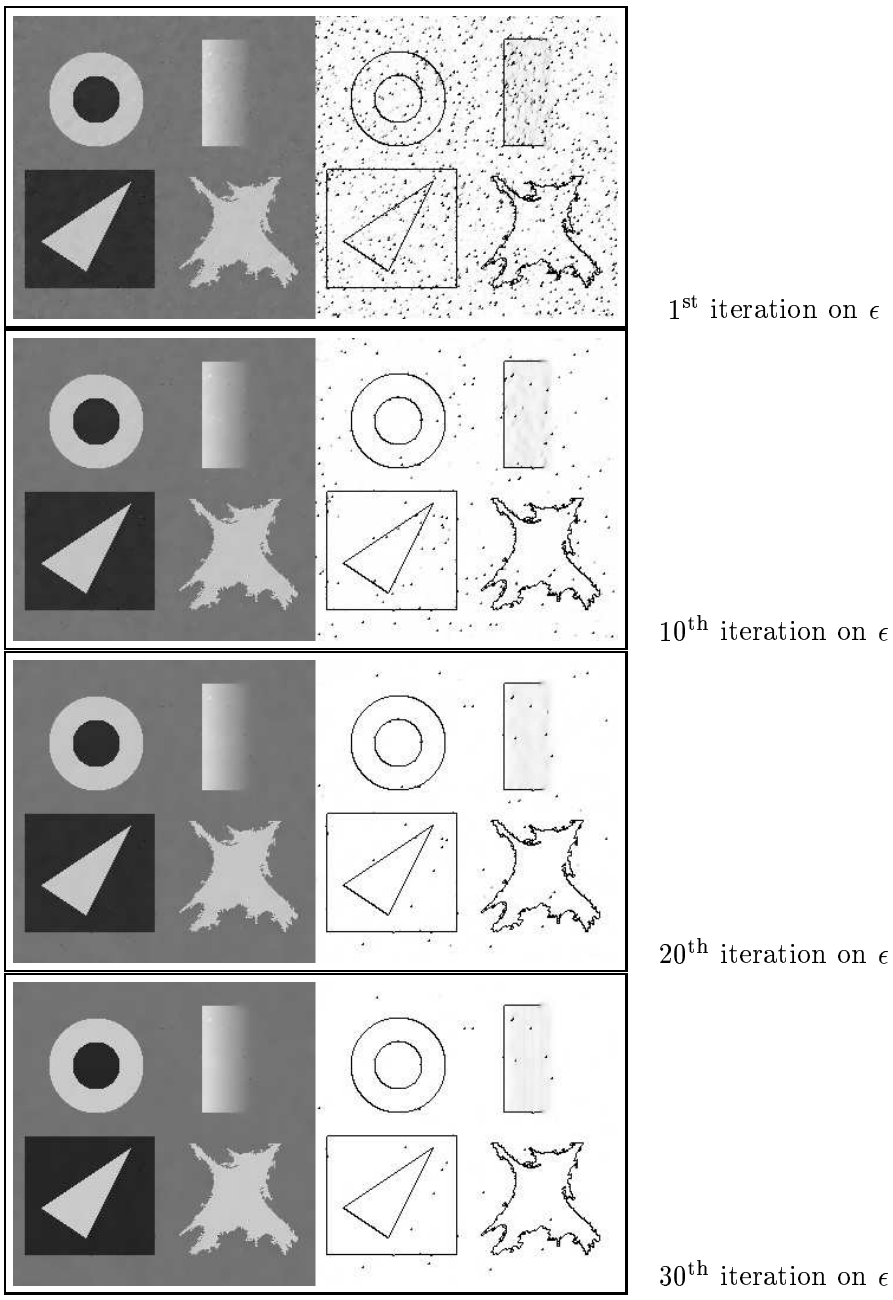
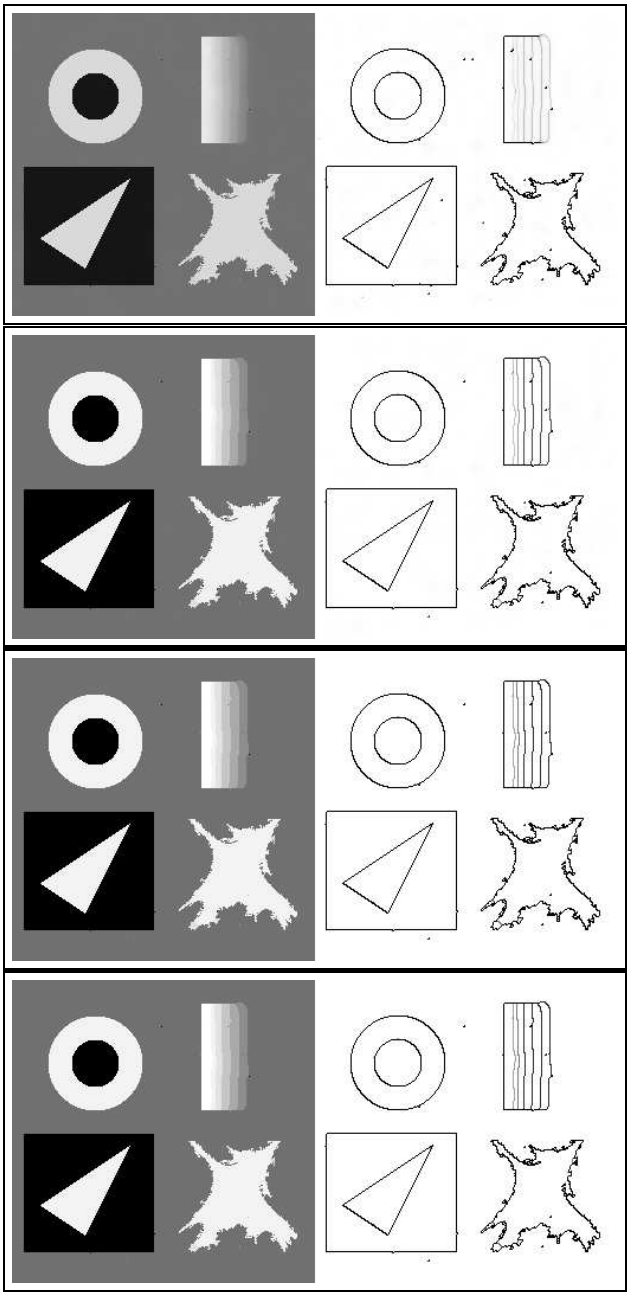


Figure 16: Solution f_ϵ (on the left handside) and associated map of discontinuities (on the right handside) for the "gdr" image with 8 classes as ϵ decreases. The higher the number of iteration on ϵ, the lowest the value of ϵ. From iteration 1 to 30: the restoration term is dominant, then after iteration 30 it vanishes whereas the level constraint term increases (see (9)).



33^{th} iteration on ϵ

36^{th} iteration on ϵ

39^{th} iteration on ϵ

50^{th} iteration on ϵ (final result)

Figure 17: Continuation of Fig. 16. From iteration 30, the value of the level constraint term is sufficiently high to let appear the classification process, i.e. the restriction of admissible values for $f(x)$ to the set of labels (it is easy to see this phenomena on the top right hanside: gradation area). The final $f_{\epsilon=50}$ result is the one presented (in color) on bottom left of Fig. 14.

5.2.2 Satellite images

We have tested the proposed classification model on SPOT images provided by courtesy of the French Space Agency (CNES).

The first 512×512 pixels image represents a Dutch agricultural area (see Fig. 18). An expert has defined ten classes, each one having an agricultural interpretation (winter wheat, permanent grass...). The parameters of the classes are presented in Table 6. The VC and the SC are shown on Fig. 19. No noise is introduced, and the regularization permits to homogenize perturbated areas.

The second image is of size 256×256 pixels (see Fig. 20) and we have defined four classes whose parameters (see Table 7) have been estimated in [10]. The VC and SC are given on Fig. 21. Fig. 22 shows the evolution of the restored image f and the map of discontinuities b while ϵ decreases. We can notice that we make enough iterations on ϵ to get a piecewise constant image f (classification). From the evolution of b we remark the progressive regularization on the boundaries: the highest the decrease w.r.t. ϵ, the sharpest the edges.

class	mean μ_i	standard deviation σ_i
1	54.6	9.65
2	73.5	2.02
3	82.5	5.96
4	93.8	9.68
5	100.5	17.57
6	122.8	2.98
7	129.9	6.11
8	146.6	3.91
9	159.9	5.59
10	182.3	8.55

Table 6: Parameters of the 10 classes for the SPOT image of a Dutch agricultural area. The means and standard deviations have been provided by the French Space Agency (CNES).

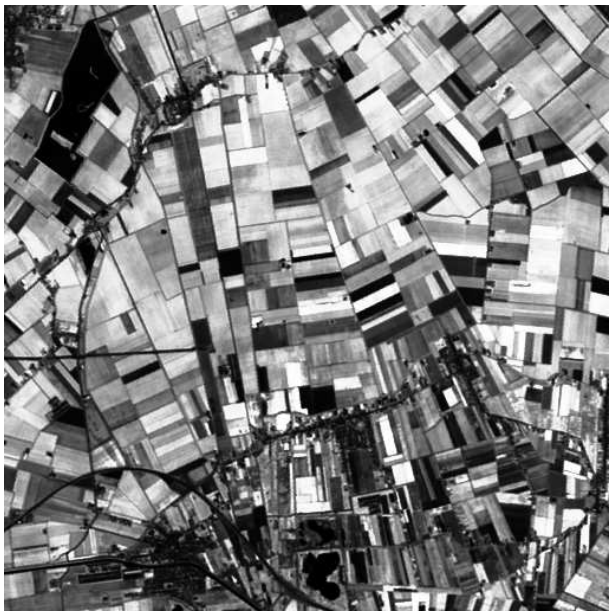
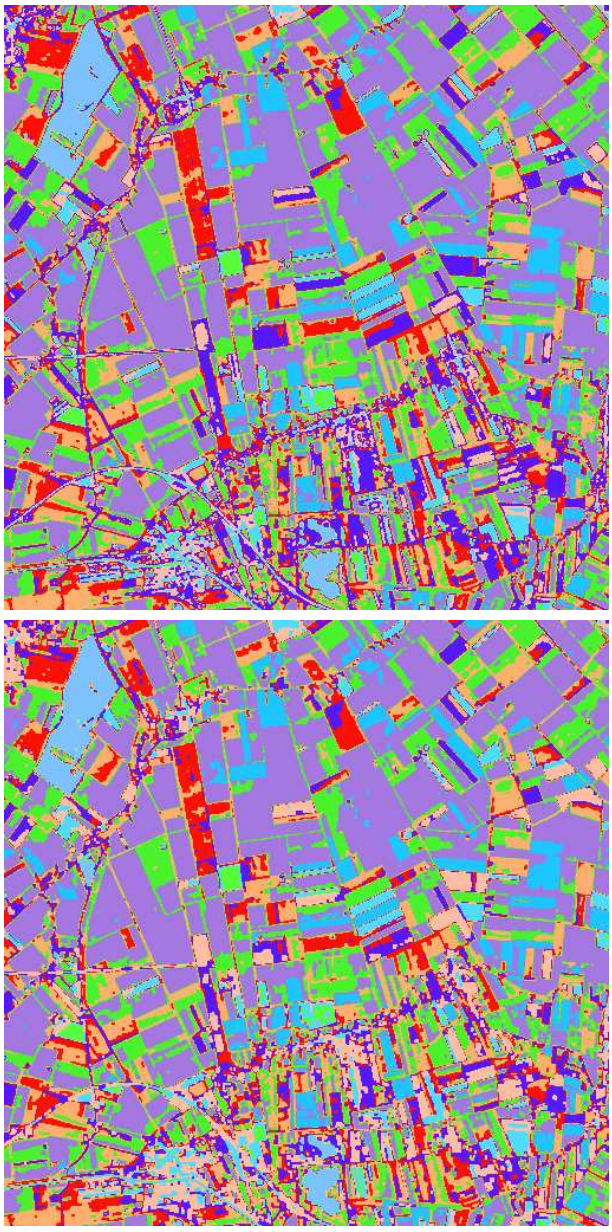


Figure 18: Original SPOT image of a Dutch agricultural area (xs3 channel, 512*512 pixels).



VC

SC

Figure 19: Results for the SPOT satellite image of a Dutch agricultural area containing 10 classes. Variational parameters: $\lambda = 6$, $\delta = 8$, $\alpha = 0.1$, 3 iterations on ϵ. Stochastic parameters: $\beta = 0.6$, 400 iterations for simulated annealing. Computational time: variational=38s., stochastic=609s.

class	mean μ_i	standard deviation σ_i
1	30.3	2.86
2	37.4	2.14
3	61.3	11.32
3	98.2	11.27

Table 7: Parameters of the 4 classes for the second SPOT image estimated in [10].

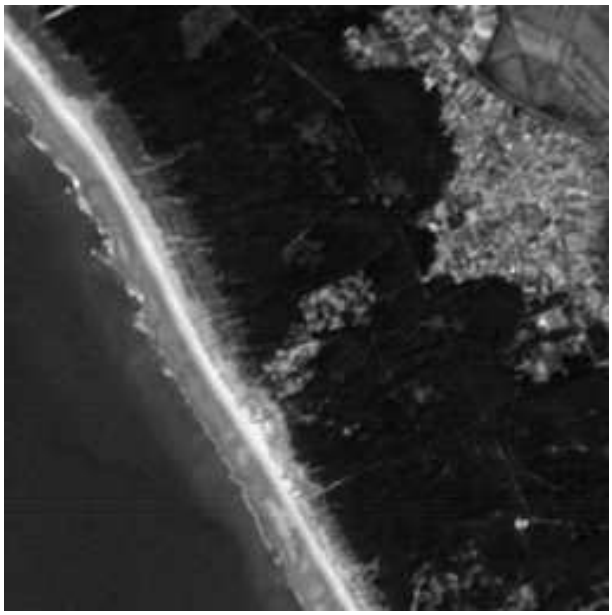
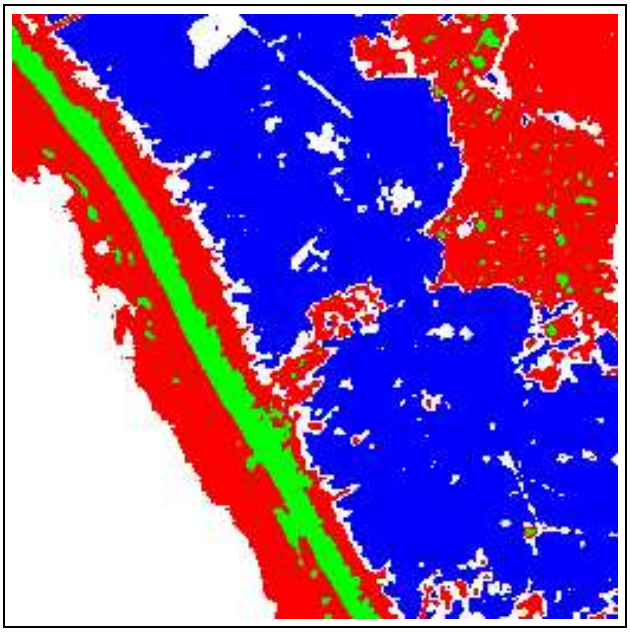
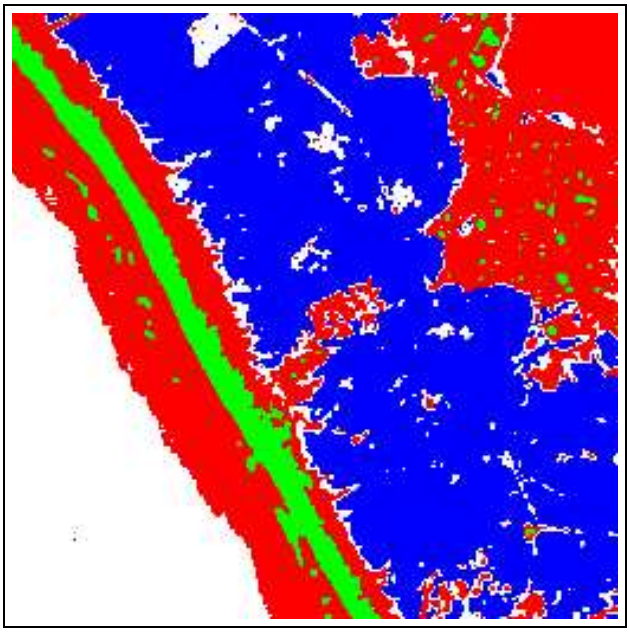


Figure 20: Second original SPOT image (256*256 pixels).



VC



SC

Figure 21: Results for the second SPOT satellite image containing 4 classes. Variational parameters: $\lambda = 5$, $\delta = 19$, $\alpha = 0.05$, 20 iterations on ϵ. Stochastic parameters: $\beta = 0.3$, 400 iterations for simulated annealing.

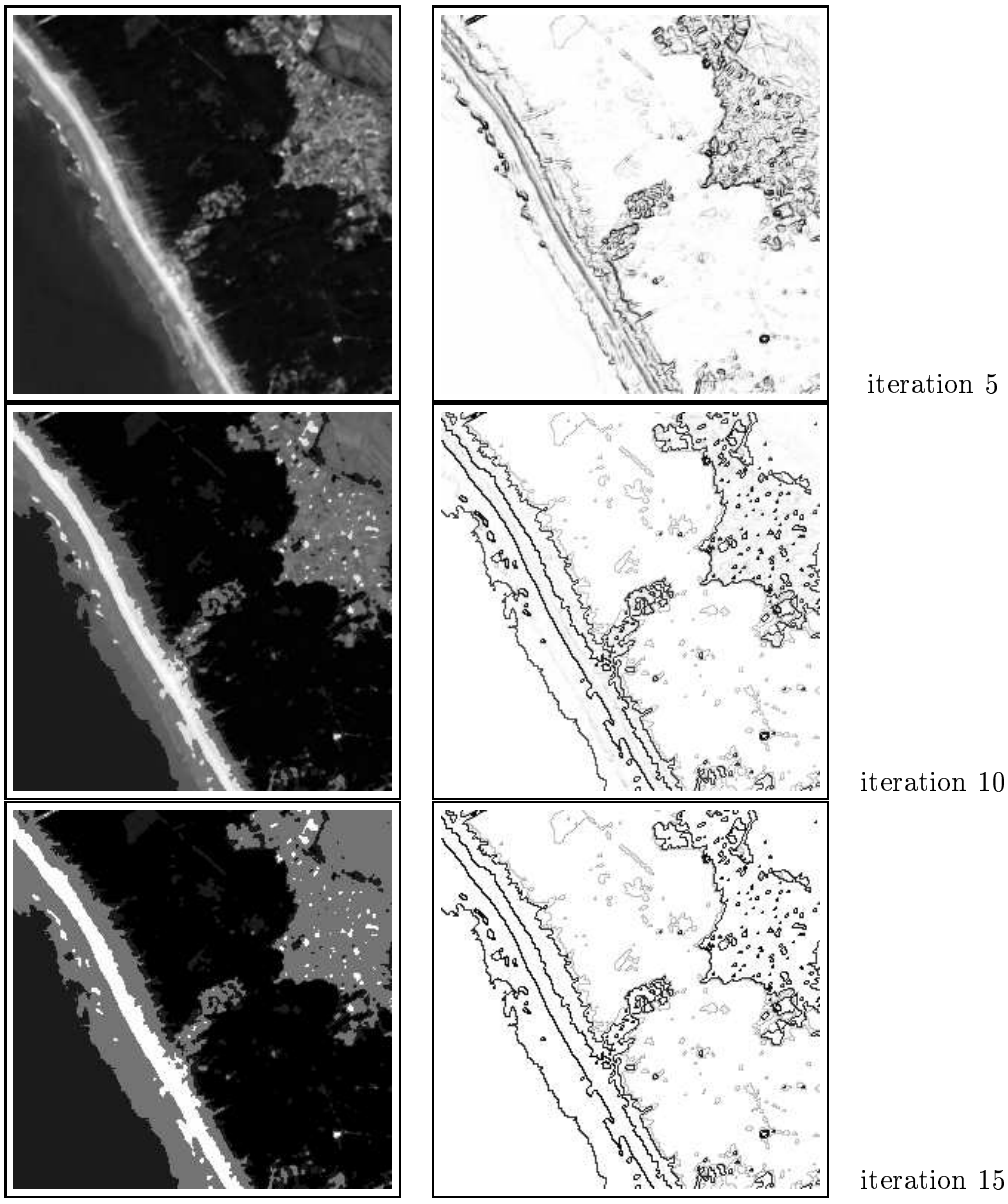


Figure 22: second SPOT containing 4 classes. Evolution of the restored image and the map of discontinuities while decreasing ϵ when increases the number of iterations.

6 Conclusion

We have presented a variational model which provides, through the minimization of a nonconvex functional, a classification and an edge-preserving restoration. This model is based on the regularization theory and mechanical phase transitions theory. Its theorical soundness holds in the case of convex regularization, but experimental results show that nonconvex regularization leads to best visual and qualitative results. The algorithm used to minimize the proposed functional is easy to implement and faster than a stochastic one for results at least qualitatively comparable. We have exhibited a delicate configuration (classes having nearly same parameters) for which the variational approach is more efficient than the stochastic model using Potts regularization. The next steps we envisage in this promising work are: first to incorporate an automatic estimation of class parameters (unsupervised classification), second we could introduce a deblurring or reconstruction process. Then, we will extend this model to the multispectral case (with direct applications to multibands satellite data and color imaging).

Aknowledgements: the authors thank the French Space Agency (CNES homepage at *http://www.cnes.fr/*) for providing the SPOT satellite images and the GdR ISIS (Working Group on Vision, Image, Signal and Information processing *http://www-isis.enst.fr/*) of CNRS for the "gdr" test image .

References

[1] S. Allen and J. Cahn. "A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening". *Acta Metallurgica*, 27:1085–1095, 1979.

[2] L. Alvarez, P.-L. Lions, and J.-M. Morel. "Image selective smoothing and edge detection by nonlinear diffusion ". *SIAM J. of Numerical Analysis*, 29(3):845–866, 1992.

[3] L. Ambrosio and V. Tortorelli. "Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence". *Commun. Pure Appl. Math.*, 43:999–1036, 1990.

[4] L. Ambrosio and V. Tortorelli. "On the approximation of functionals depending on jumps by quadratic, elliptic functionals". *Boll. Un. Mat. Ital.*, 46-B:105–123, 1992.

[5] S. Angenent and M.E. Gurtin. "Multiphase thermomechanics with interfacial structure 2. Evolution of an isothermal interface ". *Arch. Rational Mech. Anal.*, 108:333–391, 1989.

[6] G. Aubert and L. Vese. "A variational method in image recovery". *SIAM J. of Numerical Analysis*, 34(5):1948–1979, 1997.

[7] S. Baldo. "Minimal interface criterion for phase transitions in mixtures of Cahn-Hilliard fluids". *Ann. Inst. Henri Poincaré*, 7:67–90, 1990.

[8] G. Barles, L. Bronsard, and P.E. Souganidis. "Front propagation for reaction-diffusion equations of bistable type". *Ann. Inst. Henri Poincaré*, 9:479–496, 1992.

[9] G. Bellettini, M. Paolini, and C. Verdi. "Numerical minimization of geometrical type problems related to calculus of variations". *Calcolo*, 27:251–278, 1991.

[10] M. Berthod, Z. Kato, S. Yu, and J. Zerubia. "Bayesian image classification using Markov random fields". *Image and Vision Computing*, 14(4):285–293, 1996.

[11] A. Blake and A. Zisserman. *"Visual reconstruction"*. M.I.T. Press, 1987.

[12] C.A. Bouman and M. Shapiro. "A multiscale random field model for Bayesian image segmentation". *IEEE Trans. on Image Processing*, 3:162–177, March 1994.

[13] L. Bronsard and F. Reitich. "On three-phase boundary motion and the singular limit of a vector-valued Ginzburg-Landau equation ". *Arch. Rational Mech. Anal.*, 124:355–379, 1993.

[14] G. Caginalp. "Stefan and Hele-Shaw type as asymptotic limits of the phase-field equations". *Physical Review*, 39(11):5887–5896, 1989.

[15] J. W. Cahn and J. E. Hilliard. "Free energy of a nonuniform system. I. Interfacial free energy". *J. of Chemical physics*, 28(1):258–267, 1958.

[16] V. Caselles, R. Kimmel, and G. Sapiro. "Geodesic active contours". *International J. of Computer Vision*, 22(1):61–79, 1997.

[17] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. "Deterministic edge-preserving regularization in computed imaging". *IEEE Trans. on Image Processing*, 6(2):298–311, February 1997.

[18] X. Descombes, R. Morris, and J. Zerubia. "Some improvements to Bayesian image segmentation. Part one : modelling. (in french)". *Traitement du Signal*, 14(4):373–382, 1997.

[19] X. Descombes, R. Morris, and J. Zerubia. "Some improvements to Bayesian image segmentation. Part two : classification. (in french)". *Traitement du Signal*, 14(4):383–395, 1997.

[20] I. Fonseca and L. Tartar. "The gradient theory of phase transitions for systems with two potential wells". *Proc. of the Royal Society of Edinburgh*, 111A(11):89–102, 1989.

[21] S. Geman and D.-E. McCLure. "Bayesian image analysis : an application to single photon emission tomography". *Proc. Statist. Comput. Sect.*, Washington DC, Amer. Statist. Assoc.:11–18, 1985.

[22] S. Geman and G. Reynolds. "Constrained restoration and the recovery of discontinuities". *IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI 14*, 3:367–383, 1992.

[23] E. De Giorgi. "Convergence problems for functionals or operators". *Proc. of the International Meeting on Recent Methods in Nonlinear Analysis, Pitagoria, Ed. Bologna*, 1978.

[24] E. Giusti. *"Minimal surfaces and functions of bounded variation"*. Birkhäuser, 1984.

[25] P.-J. Green. "Bayesian reconstruction from emission tomography data using modified EM algorithm". *IEEE Tans. on Medical Imaging, MI 9*, 1:84–93, 1990.

[26] J. Hadamard. *"Lectures on Cauchy's problem in linear partial differential equations."*. Yale University Press, 1923.

[27] T. Hebert and R. Leahy. "A generalized EM algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors". *IEEE Tans. on Medical Imaging, MI 8*, 2:194–202, 1989.

[28] M. Kass, A. Witkin, and D. Terzopoulos. "Snakes : active contour models". *International J. of Computer Vision*, 1:321–331, 1987.

[29] Z. Kato. *"Modélisation markoviennes multirésolutions en vision par ordinateur. Application à la segmentation d'images SPOT" (in French and English)*. PhD thesis, Université de Nice-Sophia Antipolis, 1994.

[30] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi Jr. "Conformal curvature flows : from phase transitions to active vision". *Arch. Rational Mech. Anal.*, 134:275–301, 1996.

[31] S. Lakshmanan and H. Derin. "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:799–813, August 1989.

[32] R. Malladi, J.A. Sethian, and B.C. Vemuri. "Evolutionary fronts for topology independent shape modeling and recovery". In *Proc. of the 3rd ECCV*, pages 3–13, Stockholm, Sweden, 1994.

[33] B. Manjunath and R. Chellappa. "Unsupervised texture segmentation using Markov random fields models". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13:478–482, May 1991.

[34] R. March. "Visual reconstruction using variational methods". *Image and Vision Computing*, 10:30–38, 1992.

[35] R. March and M. Dozio. "A variational method for the recovery of smooth boundaries". *Image and Vision Computing*, 15:705–712, 1997.

[36] L. Modica. "The gradient theory of phase transitions and the minimal interface criterion". *Arch. Rational Mech. Anal.*, 98:123–142, 1987.

[37] J.-M. Morel and S. Solimini. *"Variational methods in image segmentation"*. Birkhäuser, 1995.

[38] D. Mumford and J. Shah. "Boundary detection by minimizing functionals". In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, 1985.

[39] C. Owen and P. Sternberg. "Nonconvex variational problems with anisotropic perturbations". *Nonlinear Analysis, Theory, Methods and Applications*, 16(7):705–719, 1991.

[40] T. Pavlidis and Y.-T. Liow. "Integrating region growing and edge detection". *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1988.

[41] F. Reitich and H.M. Soner. "Three-phase boundary motions under constant velocities. part one : The vanishing surface tension limit". *Proc. of the Royal Society of Edinburgh*, 126(A):837–865, 1996.

[42] J. Rubinstein, P. Sternberg, and J.B. Keller. "Fast reaction, slow diffusion, and curve shortening". *SIAM J. of Applied Mathematics*, 49:116–133, 1989.

[43] L. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based removal algorithm". *Physica D*, 60:259–268, 1992.

[44] P. Sternberg. "Vector-valued local minimizers of nonconvex variational problems". *J. of Mathematics*, 21:799–807, 1991.

[45] P. Sternberg and W.P. Zeimer. "Local minimisers of a three-phase partition problem with triple junctions". *Proc. of the Royal Society of Edinburgh*, 124(A):1059–1073, 1994.

[46] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud. "Variational approach for edge-preserving regularization using coupled PDE's". *IEEE Tans. on Image Processing*, 7(3):387–397, 1998.

[47] P. C. Teo, G. Sapiro, and B. A. Wandell. "Creating connected representations of cortical gray matter for functional MRI visualization". *IEEE Tans. on Medical Imaging*, 16(6):852–863, 1997.

[48] A.N. Tikhonov and V.Y. Arsenin. *"Solutions of ill-posed problems"*. Winston and Wiley, 1977.

[49] S. C. Zhu and A. Yuille. "Integrating region growing and edge detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399